A Review of Recent Advances in Learner and Skill Modeling in Intelligent Learning Environments

Michel C. Desmarais
Polytechnique Montréal

Ryan S. J. d. Baker
Worcester Polytechnic Institute

Abstract. In recent years, learner models have emerged from the research laboratory and research classrooms into the wider world. Learner models are now embedded in real world applications which can claim to have thousands, or even hundreds of thousands, of users. Probabilistic models for skill assessment are playing a key role in these advanced learning environments. In this paper, we review the learner models that have played the largest roles in the success of these learning environments, and also the latest advances in the modeling and assessment of learner skills. We conclude by discussing related advancements in modeling other key constructs such as learner motivation, emotional and attentional state, meta-cognition and self-regulated learning, group learning, and the recent movement towards open and shared learner models.

1. Introduction

It has long been recognized that individualized learning is much more effective than classroom learning. In a seminal paper, Benjamin Bloom quantified this conventional wisdom knowledge as the two sigma effect (Bloom, 1984). Based on available data, Bloom argued that the average student who received one-on-one tutoring from an expert tutor scored two standard deviations higher on standardized achievement tests than an average student who received traditional group-based instruction. Cohen, Kulik, and Kulik (1982) found similar results, though not quite the same effect size, in a meta-analysis on tutoring in general (including human tutors with less expertise than those studied in Bloom).

Achieving a similar degree of individualization has been a key interest among developers of interactive learning environments and intelligent tutoring systems. To achieve one-on-one instruction, targeted and appropriate adaptation is required, which in turn requires accurate assessment of learners.

A little over twenty years ago, John Self (1988) discussed the skepticism that many researchers held about the feasibility of an effective and useful learner model. Self argued that to overcome the difficulty of correctly assessing the learner's state and the often ovewhelming effort to build comprehensive learner models, the field should revert to didactly focused learner models of more limited scope. This view was somewhat echoed about a decade later by Cumming and McDougall (2000), who raised the question of whether Intelligent Tutoring can be main-streamed into Education. They argued that defining a strong theoretical understanding of individualized learning is a necessary condition for successully meeting this challenge. They labeled this challenge as "optimistic" for 2010.

The year 2010 is now behind us, and the 2000s have witnessed the rise of intelligent learning systems that successfully integrate learner models, and which have achieved widespread usage. To a certain extent, we can argue that the conditions proposed by Self and those of Cumming and McDougall have materialized, as these systems do incorporate didactly focused learner models and didactic strategies that yield successful one-on-one tutoring, namely what is referred to as *problem solving and solution analysis* and *curriculum sequencing* tutors.

We review these tutors, the latest developments in learner models, and the challenges that are currently being tackled. We start with the recent developments on the topic of skill modeling and assessment, which bring together the fields of cognitive modeling, psychometrics, and statistical learning.

Our review pays special attention to the techniques that have given rise to the most successful applications in intelligent learning environments. First, let us start with a short review of the basic learner model concepts.

2. Learner Models Basics

When a pupil answers that the solution to $\left(\frac{1}{3} + \frac{3}{4}\right)$ is $\frac{4}{7}$, an expert human tutor can easily observe that the pupil did $\left(\frac{1+3}{3+4}\right)$. He or she can then focus his instructions on the rule, that states that only numerators with a common denominator can be added, and can discuss how to generate equivalent fractions by multiplying the nominator and the denominator by a given factor, in order to arrive at common denominators. The tutor's ability to diagnose what a student knows and does not know, and the ability to select relevant interventions given this diagnosis, are pivotal to good tutoring. Another key feature is

tutors' ability to infer, from the student's problem solving actions and answers, what is likely well understood or mastered, and what is not, from only a few observations, and to move on in the curriculum at the right pace for that specific student. The ability to infer which skills a student masters has been, to a large extent, successfully emulated by the intelligent tutors we review in the next section. They arguably are among the most important requirements for effective one-on-one tutoring and we refer to them as *skill modeling*[1].

There are, of course, other requirements for learner models. For example, we may want to know if the student is bored or frustrated, what is the appropriate moment to switch from drill and practice to explanations and theoretical material, etc. Human tutors are well acquainted with factors like the student's attitude and motivation towards learning a given topic and their critical effect on the learning outcome (Lepper et al., 1991). In section 7 we briefly review other key issues that also play an important role in student modeling and refer the reader to a recent book by Nkambou et al. that covers many of these other factors in more details (Nkambou et al., 2010).

3. Learner Models and Learning Environments Success Stories

Increasingly, several ITSs (Intelligent Tutoring Systems) can claim to be a key part of education and learning in the real world. Their number is growing, and the most successful systems are currently used by hundreds of thousands of users a year. The fact that some systems have emerged from research labs and research classroom to widespread use is a clear indication that the models and techniques behind them have seen their potential realized. We present a few key systems, focusing on the learner modeling approaches they used and the research challenges that remain. We also focus on the systems used in public settings, as it is relatively difficult to assess the uptake of systems used only under classified settings (e.g. tutors for the military). In that respect, our review does not intend to be exhaustive, but instead to provide an overview biased towards some of the most successful and widely-used approaches to this day.

---

[1] In this paper, we define the notion of skill as encompassing problem solving abilities, concept acquisition, simple memorization of factual information and, if the tutoring context allows, even motor skills – a related conceptualization is Koedinger et al's (under review) notion of a "knowledge component" .

3.1. TUTORS FOR PROBLEM SOLVING AND SOLUTION ANALYSIS

Some of the best known success stories are from two families of tutors: Cognitive Tutors (CT) (Koedinger et al., 1997; Corbett and Anderson, 1995) and Constraint-Based Modeling (CBM) (Mitrovic, 2011). Cognitive Tutors are now distributed commercially by Carnegie Learning Inc. and reach hundreds of thousands of students each year. The Assistment system, a close cousin of Cognitive Tutors, is also gaining a strong user base (Feng et al., 2006), and is used by thousands of students a year. These systems fall into what Brusilovsky and Peylo have termed *problem solving and solution analysis tutors* (Brusilovsky, 2003). Their didactic approach is distinct from the *curriculum sequencing* approach that we review later as part of the widely used real-world learning environments.

3.1.1. *Cognitive Tutors and Constraint Based Modeling*
CT represents knowledge that is *procedural*, as it can be directly mapped to student *actions*. CBM tutors represent declarative knowledge as constraints over student *answers* (as opposed to *actions*) or over the outcome of the student's actions. Figure 1 illustrates a few examples of KT rules and CBM predicates for a problem relating to the computation of angles in an isoceles triangle.

In spite of their differences, they share strong similarities and they can achieve similar results, as demonstrated in a comparative study (Mitrovic et al., 2003).
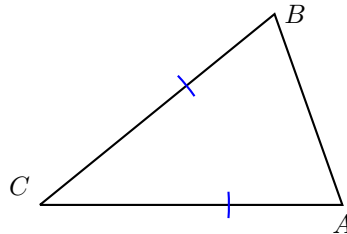
In Cognitive Tutors and Constraint Based Modeling, the skills are represented as rules (CT) and predicates (CBM), which bear a strong formal similarity. In CT, a skill is considered correctly applied by the student when a rule is matched to student performance *actions*. In the case of CBM, a skill is considered mastered when a predicate is matched over student responses.

In addition to skills, misconceptions can also be represented with these formalisms. They correspond to *buggy-rules* and to patterns of true and false predicates that reflect incorrect student knowledge. For example, figure 1's KT buggy-rule would correspond to violations of constraint 2 (that correspond to C_{r2} and C_{s2}, where C_{r2} is the relevance condition and C_{s2} is the satisfaction condition).

When incorrect knowledge is detected, this information allows immediate and fine grained remedial didactic content to be delivered to the student. This just-in-time remedial feedback is important to the student's learning and to the success of these approaches. Viewed more broadly, this feedback provides a form of scaffolding (support) that is

---

Triangle problem

Angle A is 65.
What is angle C?

Knowledge Tracing Rules

Two correct production rules:
IF goal is to find an angle in an isosceles triangle ABC and AC = BC
and angle A is known
THEN set the value of angle B to A.
IF goal is to find an angle in a triangle ABC and angles A and B are known
THEN set the value of C to 180-A-B
Buggy production rule:
IF goal is to find an angle in an isosceles triangle ABC
and angle A and C are at the bottom of the triangle and angle A is known
THEN set the value of angle C to A.

CBM Constraints

C_{r1}: A base angle of an isoceles triangle is known (θ_1), and the student has calculated the size of the other base angle (θ_2).

C_{s1}: The size of θ_2 is θ_1.

C_{r2}: A base angle of an isoceles triangle is known (θ_1), and the student has calculated the size of another angle θ_2 that equals θ_1.

C_{s2}: θ_2 is a base angle.

C_{r3}: Two angles of a triangle are known. (θ_1 and θ_2), and the student has calculated the size of a third angle θ_3.

C_{s3}: The size of θ_3 is $(180 - \theta_1 - \theta_2)$.

Figure 1. KT rules and CBM constraint examples (adapted from Mitrovic et al., 2003).

delivered at the appropriate time, in order to be most effective (Lajoie, 2005).

3.1.2. *The Role of Student Models in CT and CBM*

It is interesting to note that the scaffolding and just-in-time remedial feedback that is critical to CT and CBM tutors does not necessarily imply any substantial "long term" student model. A tutor may solely track the user actions from the last problem, and the successful and unsuccessful attempts within the current topic, to decide whether to move on to the next topic. Then, it can start anew given this next topic, as if it were dealing with another student. This student would still receive hints and remedial content based on the cognitive diagnosis of his or her actions and answers. This analysis of user actions to track the problem resolution state is termed *Model Tracing* and it bears resemblance to plan recognition techniques.

Model Tracing allows the tutor to give feedback and hints, akin to the process of identifying which constraints are satisfied or not in CBM. However, as the skills assessed by Model Tracing typically span multiple topics and exercises, it is desirable to maintain a student model that provides an accurate assessment of all relevant skills, over time. In Cognitive Tutors, this is known as *Knowledge Tracing* and is now used in most CT (Corbett and Anderson, 1995). We return to Knowledge Tracing in section 5.4. In CBM tutors, similar techniques have been devised, some based on overlay models and others based on probabilistic approaches (see Mayo and Mitrovic, 2001, and Mitrovic, 2011).

3.2. CONTENT SEQUENCING TUTORS

Another family of tutors that have enjoyed substantial success today are environments that guide a student through learning material. The most widely used is probably the ALEKS system (www.aleks.com), which is now said to be used by millions of users. ALEKS is a commercial spin off of the University of California at Irvine and is based on the cognitive theory of *knowledge spaces* (Doignon and Falmagne, 1999; Doignon and Falmagne, 1985). This theory is at the basis of a number of efforts and active developments in the field of learner modeling (Heller et al., 2006; Desmarais et al., 2006).

The ALEKS tutor takes a very different approach to tutoring. This approach can be considered *curriculum sequencing*, a concept that can be traced back to McCalla, Peachey and Ward (1982) and Peachey and McCalla (1986), and which consists in defining learning paths in a

space of learning objectives and didactic content (see also Brusilovsky and Vassileva 2003, and Vassileva, 1995).

Whereas CT and CBM aim to provide specific remedial content based on a detailed analysis of the student's problem solving steps or answers, curriculum sequencing aims to make broader skills assessment to adapt the learning content in general. Adaptive hypermedia (Brusilovsky, 2001), used as a learning tool, is a representative example of this approach. Adaptation can be as coarse-grained as recommending courses or book chapters, or as fine-grained as choosing exercises deemed new and challenging, yet not too difficult, in accordance to Vigotsky's *zone of proximal development.* (Vygotsky, 1978). Other systems like SIETTE (Conejo et al., 2004) (limited to test items sequencing) and RATH (Hockemeyer et al., 1997) (also based on the theory of knowledge spaces) adopt a curriculum sequencing approach. See also Brusilovsky and Peylo (2003), who have reviewed a number of similar approaches.

Whereas the success of problem solving and solution analysis tutors, such as CT and CBM, relies on the ability to provide just-in-time remedial feedback and decide when to move on to a new topic, the success of curriculum sequencing lies in tailoring the learning content based on an accurate assessment of a large array of skills with the least possible amount of evidence. The ratio of the amount of evidence to the breadth of the assessment is particularly critical for systems that cover a large array of skills, as it would be unacceptable to ask hours of questions before making a usable assessment. This requires a model that can build links among skills, such as prerequisites; for example, such a model would infer that a student's knowledge of English vocabulary terms such as "sibilate" and "quandary" clearly indicates that familiar common terms would probably be unchallenging. Modeling prerequisite structure is not as critical for tutors like CT and CBM, that adopt a problem solving and solution analysis approach. As a consequence, learner models for CT and CBM tutors on one side, and for content sequencing tutors on the other side, have diverging emphasis which lead them to adopt different (though not incompatible) approaches.

4. Learning Models Revisited

In retrospect, we can argue that the early success of intelligent learning environments came from two sources. One of them is the support for highly specific, immediate, and effective feedback during problem solving. The value of this feature to the student was prominent in CT (Cognitive Tutors) and in CBM (Constraint Based Modeling) tutors.

Another source was the ability to structure the learning path according to the individual skill profile of each student. This feature was the most salient in ALEKS, and it required that the learning application be able to define meaningful learning paths and to build an accurate and global assessment of the student skills with only partial evidence, from a few questions.

The first source of success can be considered as fine-grained skill diagnosis that relies on a detailed domain analysis of each exercise proposed to the student, whereas the second source relies on what has been called a *transfer model*: a model that can perform a multi-skill assessment from a subset of observed skill mastery or, in other words, a model that transfers evidence between skills, or between items and skills that are not directly linked[2].

The distinction between these sources can be conceptualized according to Figure 2's layers. The learner actions are displayed at the lowest level of the network. One or more actions can be directly mapped onto nodes in a model. A sequence of actions can therefore trigger the activation of a node in the observable nodes layer, and indicate mastery of a skill, or it can trigger a *mal-rule* indicating a misconception.

Moreover, one or more actions can also generate an outcome (second layer up in Figure 2). The outcome can change the problem state which, in turn, can be used as a constraint in a CBM tutor. In that respect, the Action outcome layer is one level of abstraction above the actions, since many different action sequences can result in the same outcome. Akin to mapping actions to one or more specific observable skills in the model, Outcome nodes can be mapped to one or many skills nodes in the model.

A tutor that does not transfer knowledge of what is considered mastered between exercises would have its model nodes tied to observable actions only. Obviously, for highly effective transfer, the model's nodes need to be fully connected. There are many ways to achieve node connection. Models behind curriculum sequencing such as ALEKS build links among observable nodes themselves, without hidden nodes. Another approach, Item Response Theory which we review later, links every observable node to a single hidden node. Finally, Bayesian Networks can contain many layers of hidden nodes, albeit by imposing constraints or assumptions to allow the calibration of links among hidden nodes that, otherwise, would be very error prone without any empirical input from direct observations (see Reye, 2004).

----

[2] The term *transfer model* has been used in cognition and learning to express different phenomena. We use it here in a sense close to that of Pavlik et al. (2009).
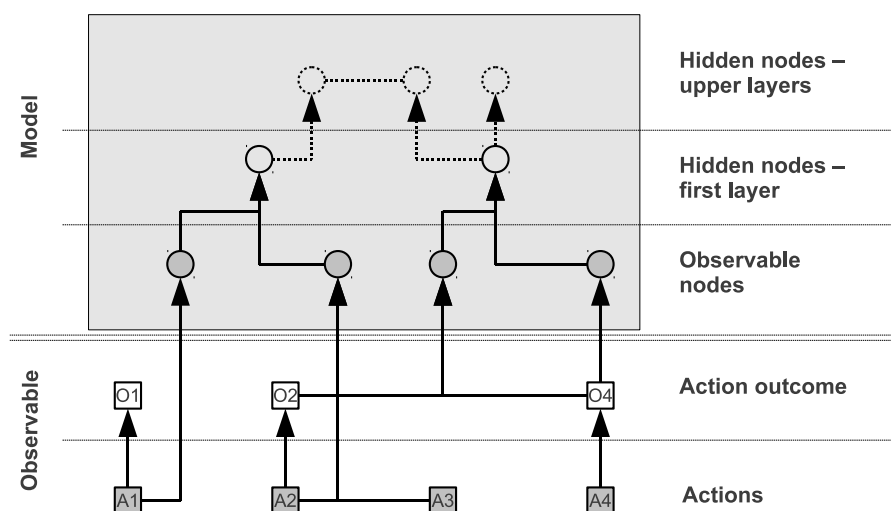
Figure 2. Learner modeling layers.

Not represented in Figure 2 is the temporal dimension of observations and the fact that a node probability changes over time, along sequences of observations. These types of models have been the focus of intense research, in particular following the seminal work of Corbett and Anderson (1995). They are temporal bayesian networks (Dynamic Bayesian Networks) that share resemblance to Markov Models (see section 5.4).

In the next section, we go into more details describing the models we mention above.

5. Models of Skills

Uncertainty is a factor that must be dealt with when modeling and assessing skills. For example, after how many successful opportunities should we consider a skill mastered? The answer depends on many factors. The occurrence of *slips*, when a student accidentally (or due to carelessness) fails a known item, and *guesses*, when the student correctly answers an item by chance, are important sources of uncertainty (Baker and Corbett, 2008). Uncertainty is particularly important for transfer models when the goal is to build global assessment from indirect evidence. For example, if a pupil succeeds at solving $\sqrt{(2^6)} = 8$, a transfer model may conclude that this student would also solve $3/4 \times 2/5$. But because the specific skills involved are different, requiring transfer of evidence from one problem to the other, we consider this evidence as indirect and inherently uncertain.

We review the major modern probabilistic approaches to skill modeling within the light of these considerations. We start with the most general approach, Bayesian Networks.

5.1. Bayesian Networks and Graphical Models

A Bayesian Network (BN) is a highly flexible graphical and probabilistic modeling framework that has the potential to encompass all modeling layers found in figure 2. BNs have been widely used for the purpose of modeling learner skills. Among the best known are the ANDES Assessor (Conati et al., 2002; Vanlehn et al., 2005), HY-DRIVE (Mislevy and Gitomer, 1995), and Zapata-Rivera and Greer's (2003) inspectable student models. The attractiveness of Bayesian models comes from their high representative power and the fact that they lend themselves to an intuitive graphical representation. In addition, Bayesian Networks (BN) offer a well defined formalism that lends itself to sound probability computations of unobserved nodes from evidence of observed nodes. A number of software applications and libraries are available to compute probabilities based on observed nodes. Moreover, BNs can potentially be derived from data, thereby reducing the need for substantial knowledge engineering (see Neapolitan, 2004). However, in practice, issues arise when we need to determine the topology and probability parameters of hidden nodes. We already hinted on the problem of parameter learning in BN and will revisit it again as it is a critical issue.

Figure 3 illustrates a potential Bayesian network in the domain of fraction arithmetic. Skills or concepts involved (top) are linked to specific items to solve (bottom). Some items involve a single concept, whereas others can involve two and potentially more. In addition to representing skills, misconceptions can also be combined within the BN framework and they can be linked to skills and items. Furthermore, skills themselves can be linked among themselves to express higher order knowledge ($C1 \rightarrow C2$ and $C1 \rightarrow C3$). A gentle introduction to the topic can be found in Jameson (1995) and a more recent surveys can be found in Mayo and Mitrovic (2001), Almond et al. (2007) and Reye (2004), and in recent books by Woolf (2009) and Nkambou et al. (2010).

Links represent interdependencies that are captured in the form of Conditional Probability Tables of a child node given its parents ($P(\text{child} \mid \text{parents})$). Following the semantics of Bayesian Networks, we would conclude, from figure 3's network, that if we find out that a student knows C2, then that would increase our belief that C3 is also
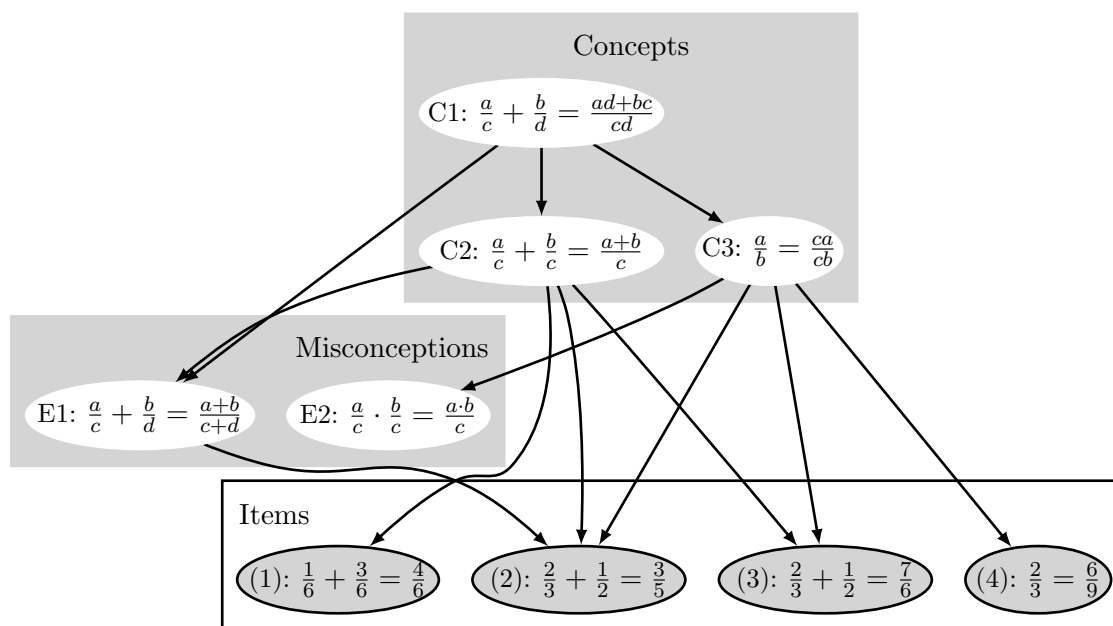
Figure 3. Hypothetical Bayesian Network on fraction arithmetics. Concept and mis-conception are *hidden nodes* that represent algebraic transformation rules that a student can apply in solving exercises, whereas items are observable nodes representing exercises (left-hand-side of equality) and student responses (right-hand-side).

known[3], but if we knew that C1 is also known, then further discovering that C2 is known would not affect our belief about C3. This property is known as the Markov Blanket and it is a fundamental concept of BN. Not only does the topology of BN inform us of which nodes are parents and children, it also defines the conditional independence between nodes. Returning to the example above, figure 3 BN informs us that C2 and C3 are conditionnally independent given C1.

Now, the links between C1, C2, and C3 need not be cognitively nor conceptually meaningful. In this particular case, the links are based on the principle that the decomposition rule C1 would generally imply that the student also knows about the decomposition rules C2 and C3, since they are involved in the formal derivation of rule C1. This principle is congruent with the Markov Blanket condition. In general, such kinds of logical or practical relations will indeed reflect the interdependencies between concepts, misconceptions, and items. Nevertheless, it should

---

[3] Or, alternatively, decrease our belief that C3 is known. The Bayesian Network formalism does not prescribe increase or decrease in probability. It solely defines conditional dependencies and independencies.

be stressed that it may not represent the optimal network topology for predictive (skill assessment) purposes.

Furthermore, attemps to guide the definition of a Bayesian network based on sound cognitive or conceptual principles may not be compatible with the objective of delivering an optimal probabilistic skills assessment model. For example, although the granularity hierarchies of Greer and MCalla (1989; McCalla and Greer, 1992) form a sound basis for defining efficient skills abstractions for diagnosis using constraint satisfaction and plan recognition techniques, their semantics may clash with the Bayesian network's own semantics.

5.1.1. *Issues and Challenges*

Bayes Nets have been popular in the field in large part due to their combination of (1) flexibility (any node can, in theory, be given as new evidence and all other node probabilities will be updated accordingly), (2) high expressiveness in a unifying framework (concepts, items, and misconceptions all follow the same semantics for computation of probabilities), and (3) sound computations that can be carried out by standard software packages.

However, these advantages come at a price. The most challenging issue is that such networks can contain a high number of hidden nodes, such as concepts/skills and misconceptions. Because hidden nodes are never directly observed, deriving the conditional probability tables from data can prove very complex as it becomes subject to the so-called *curse of dimensionality* (see Hastie et al., 2001).

Vomlel (2004) addressed the issue of setting the conditional probability tables by reverting to a combination of expert derived estimates and a data driven approach. Using standard algorithms to derive the structure of a BN from data, he used this information to validate the structure of relations among skills in fraction arithmetic that are similar to figure 3 and encompass a total of 20 hidden nodes. By asking experts to assess whether each of 149 student mastered the 20 concepts from their test results, he was able to derive part of the BN structure from this data and refine the structure with experts. The conditional probability tables were also derived in this manner.

However, the process of requiring experts to assess concept mastery of students from test results is in general impractical and prone to errors. Thus, the majority of systems that use a BN learner model revert to simplifying assumptions in order to determine conditional probabilities (Conati et al., 2002; Carmona et al., 2005; Millán and Pérez-de-la-Cruz, 2002; Almond et al., 2001; Mislevy et al., 1999). Some of the key simplifying assumptions used in learning systems are discussed in the following sections.

5.1.2. *Noisy-AND, Leaky-OR Gates Assumptions*

One such simplifying assumption is the notion of a Leaky-OR Gate, where, in the context of a set of an item or a concept having many parents, any parent is considered sufficient to succeed the item child, or to master the concept child. An alternative assumption is the Noisy-AND Gate where all parents are required be true in order for the child to be true.

Under the Noisy-AND gate assumption (VanLehn et al., 1998), the probability of a correct response to a child node X (assuming it is an observable test item) given its parents (i.e. the skills involved to solve the item) is:

$$P(X \mid \text{all parents} = \text{T}) = 1 - slip$$
$$P(X \mid \text{at least one parents} = \text{F}) = guess$$

In the case of the Leaky-OR gate, probabilities are obtained by:

$$P(X \mid \text{one or more parent} = \text{T}) = 1 - slip$$
$$P(X \mid \text{all parents} = \text{F}) = guess$$

The *slip* and *guess* parameters are generally global to all items and tasks, although they can be set according to rules such as *guess/number of answers to that problem* (VanLehn et al., 1998), in which case it can be different on a per-item basis.

Versions of the Noisy-OR gate approach was adopted by Conati et al. (2002) for the Andes system and followed by Carmona et al. (2005), among others. These approaches can be considered simplified versions of the NIDA/DINA (Noisy-AND) and NIDO/NIDO (Noisy-OR) models, that were developed within the psychometrics field and which are reviewed in section 5.3. The difference lies in the number of individual parameters that each model handles, namely if individual guess and slip parameters are defined for each skill and each item.

5.1.3. *Data Driven Bayesian Networks*

Instead of using simplified assumptions, another approach consists in dealing only with observable nodes and limiting the size of the network. For example, Mayo and Mitrovic (2001) used a structure learning algorithm with training data to identify the most relevant tasks and events (constraint violation) related to the current task and parametrize a BN with these nodes only. Then, based on the previous success and failures of the student, the BN is updated to assess the chances of success to the current task.

A different BN is derived for each problem, which makes the approach genuinely dynamic. The constraint of building a BN with observable nodes only make the BN induction and parametrization tractable

in practice. They term their approach *data-centric* because models are derived from data as opposed to domain expert engineered models (*expert-centric*) or approaches that rely on strong simplifying assumptions (*efficiency-centric*). The principle of constraining the model to observable nodes only is a key factor to the success of this approach and has been followed by many other approaches we review below.

5.2. IRT AND LATENT TRAIT MODELS

IRT (Item Response Theory) is a prominent approach that has been studied in the field of psychometics for over 40 years and formed the basis of the first personalized assessment environments, Computer Adaptive Tests. Only recently have researchers from this community and the ITS community begun to draw from the other tradition's rich history.

As a learner skills model, IRT can be considered as the ultimate transfer model. Referring to figure 2 layers, IRT would contain a single hidden node (the *latent trait*) that is linked to all observable outcomes, where the outcomes are the success or failure to test items. Any observation of an outcome therefore leads to a transfer of evidence to all other items throught the latent trait.

The fundamental idea behind IRT is that the chances of success on a task or, to use the standard IRT terminology, to an *item*, increases as a function of the level of mastery of a latent, unobserved skill, θ. According to the psychometric theory, the shape of this function corresponds to the integration of a normal distribution (cumulative distribution function) and it is termed the Item Characteristic Curve (ICC), which can be closely approximated by the logistic distribution. Therefore, IRT can be conceived as a logistic regression model where, given a vector of responses $\mathbf{X}$, skill mastery, θ, is estimated by maximing the equation:

$$P(\theta|\mathbf{X}) = \prod_i P(X_i|\theta)$$

As noted by Almond et al. (1999), this is similar to a Naive Bayes structure where θ is the root node of a set of item nodes, $\mathbf{X}$, except that instead of having each $P(X_i|\theta)$ modeled as a conditional probability table, it is modeled as a logistic function

$$P(X_i|\theta) = \frac{1}{(1 + e^{-a_i(\theta - b_i)})} \tag{1}$$

where a corresponds to the item's discrimination power and b corresponds to its difficulty.

In its standard form, IRT is a single skill model, which makes it unfit for fine-grained cognitive diagnosis. However, it enjoys a strong

theoretical background both in terms of being grounded in psychometric measurement, and a sound mathematical framework with proven algorithms (Baker and Kim, 2004).

There are variants called multidimensional-IRT that can handle two and more dimensions (Briggs and Wilson, 2003; Reckase and McKinley, 1991), but their complexity is much greater, and they have not yet been used widely in personalized learning environments.

One of the earliest and most complete effort to embed IRT as a cognitive diagnostic tool in a system that can yield a fine-grained cognitive diagnosis was conducted by Millàn and Pérez-de-la-Cruz (2002). They integrated the principles of IRT with a hiearchical structure of concepts to derive a detailed assessment. The probability of mastery of a concept increases as the number of its children are considered mastered, and the link function is inspired from IRT's ICC curve. Another similar effort has been developed by Guzman et al. (2007) and integrated into the SIETTE adaptive testing system (Conejo et al., 2004).

5.3. LATENT COUSINS DINA, NIDA, DINO, NIDO

Developments in the field of psychometrics have brought a class of latent models that can be considered as generalization of the AND/OR Gate models we have already encountered in section 5.1.2 (see Roussos, 2007, for an overview of these models; and Junker and Sijtsma, 2001). They are based on the notion that each task is linked to a number of skills (dimensions), akin to the notion of *opportunities* to practice a skill for given task in Knowledge tracing. The mapping between tasks and skills is represented by a Q-matrix (Tatsuoka, 1983) which defines the links between items and skills. Assuming we have I items and K skills, then the Q-matrix is defined as:

$$\text{items} \begin{array}{c} \text{skills} \\ \begin{pmatrix} q_{1,1} & \cdots & q_{1,K} \\ \vdots & \ddots & \vdots \\ q_{I,1} & \cdots & q_{I,K} \end{pmatrix} \end{array}$$

For example, if an item x_1 involves only skills k_2 and k_3, then $q_{1,2}$ and $q_{1,3}$ will be set to 1, and all other entries for that item, $q_{1,\bullet}$, will be set to 0. Q-matrices are in fact considered a form of *transfer model* which can link items to concepts, or even concepts together (see for eg. (Pavlik et al., 2009a)).

When an item involves multiple skills and when the low mastery of a single one of them is sufficient for failing this item, the model is considered part of the *conjunctive* class, signifying that all skills are necessary to succeed the corresponding item. Conversely, if a strong

mastery of a single skill is sufficient to succeed at the item, it will be considered part of the *compensatory* class of models.

The NIDA model (Noisy Input Deterministic And) (Junker and Sijtsma, 2001) is a conjunctive model: All skills involved in an item must be mastered to succeed. It also makes the assumption that all items within a skill have the same *guess* and *slip* parameters. However, *guess* and *slip* vary across skills.

The probability of a correct response to an item X is $1 - s_k$ (*slip*) if all skills involved are mastered and g_k if any skill is not mastered (*guess*). The assumption that the *slip* and *guess* parameters vary across skills but not across items renders their estimation feasible even with very small samples, but these assumptions are obviously unrealistic in many contexts.

The NIDO model is the NIDA model's disjunctive (compensatory) counterpart: it assumes that a single skill is necessary to succeed the item. The probability of a correct response is 1 minus s_k if any skill involved is mastered and g_k if no single skill is mastered.

The DINA model (Deterministic Input Noisy And) (Haertel, 1989) associates the *guess* and *slip* parameters to items instead of skills. It makes the same underlying assumption as the NIDA model, namely that all required skills must be mastered for the item to be succeeded. But different *guess* and *slip* parameters are associated with each item. The *guess* parameter represents the chances of success given nonmastery of one or more skills. Conversely the *slip* parameter represents the chances of failure given mastery of all items. Akin to NIDO, the DINO model is the compensatory counterpart to the DINA model. Because these models comprise a large number of parameters, they require more data for their estimation. Yet, they make no distinction for the number or the nature of the skills involved for an item, which also is unrealistic in many contexts.

By dropping some assumptions and by introducing different parameters and assumptions, a number of variants to these models can be introduced. We refer the reader to Roussos et al. (2007) and Junker and Sijtsma (2001) for more details. Futher details about compensatory and non compensatory models with continuous estimates of skills are reviewed in Stout (2007).

5.4. Bayesian Knowledge-Tracing

Bayesian Knowledge-Tracing (BKT) is another approach that relies on Bayesian theory. It operates at the level of learner actions and observable nodes (see figure 2), but it integrates a notion of *time sequences*. The approaches reviewed so far assume a static learner knowledge state,

whereas BKT models learning in time. This approach to skills modeling is particularly relevant for tutors that use exercises and scaffolding as the main vehicle for learning and that monitor fine-grained skill mastery to decide on the next step.

The BKT technique is used in Cognitive Tutors (Corbett and Anderson, 1995) and it has gained widespread acceptance. Bayesian Knowledge-Tracing continues to be the subject of intensive research, focused on ways to improve upon the base model without losing the simplicity and tractability that characterizes BKT.

BKT is essentially a model for determining if and when the learning of a skill (or other type of knowledge component) occurs during a specific problem-solving step. Assuming that each step of each learning exercise calls for a given single skill, an opportunity to demonstrate (and learn) that skill occurs and the student can either succeed or fail the task. In the basic model, four parameters are defined:

$P(L_0)$**:** Probability the skill is already mastered before the first opportunity to use the skill in problem solving.

$P(T)$**:** Probability the skill will be learned at each opportunity to use the skill.

g**:** Probability the student will guess correctly if the skill is not mastered (*guess*).

s**:** Probability the student will slip (make a mistake) if the skill is mastered (*slip*).

$X_n = \{1, 0, x_n\}$**:** {Correct outcome, Incorrect outcome, Outcome$_n$} for the item corresponding to opportunity n to use the skill.

The probability that the skill L at opportunity n is mastered can be computed as:

$$P(L_{n-1}|X_n=1) = \frac{P(L_{n-1})(1-s)}{P(L_{n-1})(1-s) + (1-P(L_{n-1}))g}$$

$$P(L_{n-1}|X_n=0) = \frac{P(L_{n-1})s}{P(L_{n-1})s + (1-P(L_{n-1}))(1-g)}$$

$$P(L_n|X_n=x_n) = P(L_{n-1}|X_n=x_n) + (1 - P(L_{n-1}|X_n=x_n))P(T)$$

The BKT model can be considered as a Markov Model to the extent that the probabilities above depend either on fixed parameters and on the previous state, $n-1$. BKT models can also be considered simple Dynamic Bayesian Networks, which are a special class of Bayesian Networks for representing temporal information[4]. The particular topology of such networks allows for sound and tractable computations of the predictions as well for the learning of parameters from data. There

---

[4] Dynamic Bayesian Networks are also considered a generalisation of single-state Hidden Markov Models

exists ongoing debate as to the best approach for computing parameters, including Expectation Maximization and Brute Force/Grid Search (Pavlik et al., 2009b; Gong et al., 2010).

Recent work in extending Bayesian Knowledge-Tracing has introduced a number of advances, such as the contextualization of estimates of guessing and slipping parameters (Baker et al., 2008a), estimates of the probability of transition from use of help features (Beck et al., 2008), and estimates of the initial probability that the student knows the skill (Pardos and Heffernan, 2010a). Item difficulty has also been recently integrated in this model (Pardos and Heffernan, 2011).

Performance Factors Analysis (Pavlik et al., 2009b) and Learning Factors Analysis (Cen et al., 2006; Pavlik et al., 2009a) are alternative approaches to BKT for inferring changes in knowledge from sequential data that may offer advantages over BKT (see also Yudelson et al., 2011). Yet another alternative proposes to use a logistic regression based approach which is particularly efficient in the case of tasks involving multiple skills (Xu and Mostow, 2011).

The BKT approach is a sound and well defined scheme to assess the observable node layer. However, to perform a global skills assessment, typical of curriculum sequencing approaches, it needs to be complemented with a transfer model. The Q-matrix is one such model that can serve as a complement (see section 5.3), and efforts to conduct such an integration have begun (Pavlik et al., 2009a).

5.5. MODELS WITHOUT HIDDEN NODES

The last class of models that we consider are based on the theory of *Knowledge spaces*, where a learner's state of knowledge is represented as a subset of items representing *knowledge units*. The distinctive characteristic of models based on knowledge spaces is that they contain only observable nodes (refer to figure 2).

Representing a student's knowledge state as a subset of observable nodes is similar to the well-known *overlay* approach used in ITS, with the exception that items correspond to observable task outcomes instead of concepts/skills. Furthermore, the theory indicates which knowledge states can be reached from a given knowledge state, based on surmise relations among items. Surmise relations impose an order among items. This order represents the constraint that a student learns to solve simpler problems before moving on to more complex ones. To demonstrate this principle of item orders, let us reuse items (1), (3), and (4) of figure 3, and add a new item, (5). Figure 4 illustrates the likely surmise relations between these items. Surmise relations are similar to logical implication relations and do not convey the same semantics as a
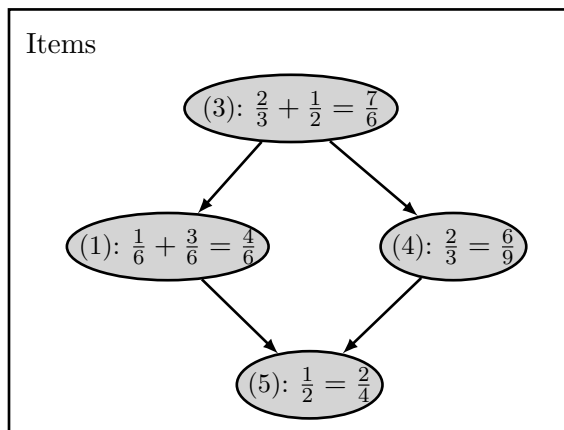
Figure 4. Example of a knowledge structure. Each node represents an observable test item where the answer is given on the right-hand-side. The links indicates the order in which a typical student would learn to solve the problems. The corresponding knowledge space is $\{\{3, 1, 4, 5\}, \{1, 4, 5\}, \{1, 5\}, \{4, 5\}, \{5\}, \emptyset\}$ and it represents all valid knowledge states.

Bayesian Network. The relations stipulate that if a student gets item (1) or item (4) right, we can infer that item (5) will also be succeeded on. Conversely, if the student fails item (4), then item (3) is likely to be failed on also. This structure dictates there is no clear order between items (1) and (4).

As mentioned, knowledge strutures do not contain concepts, nor misconceptions, which would be represented by hidden nodes. However, it is possible to derive their structure from data and recover subtle relations between items such as the fact that item (4) is more difficult than item (5) for most pupils (because of the actual numbers involved), even though the two items involve the same algebraic concepts (C3 in figure 3). To go from items to concepts, one could use a Q-Matrix where each item is linked to concepts, possibly with weights that indicate the relative importance (akin to scoring items of an exam). In addition to this simple scheme, other approaches have been developed that extend the knowledge space framework to include skills (eg. Heller et al., 2006; Dütsch and Gediga, 1995).

The simplest approach to modeling based on observable nodes is the POKS model (Partial Order Knowledge Structures). POKS was originally introduced in (Desmarais et al., 1996) and a refined version was reported in (Desmarais et al., 2006). The approach consists in deriving a partial order among item similar to figure 4. This formalism is a special case in the Knowledge Spaces theory, because it does not

allow alternative paths: if an item, say an exercise, involves two alternate methods to solve it, then only one of the children node must be mastered, as opposed to imposing that all children nodes be mastered as in figure 4. The formalism to represent knowledge structures in the Knowledge Spaces theory is known as an AND/OR graph instead of a partial order. The partial order simplification greatly reduces the amount of data that is required to induce the knowledge structure.

Other approaches that derive from the theory of Knowledge spaces can be found as early as 1992 (Villano, 1992), as well as more recently in the theory of Competence-based Knowledge Space (Heller et al., 2006). An active line of research is to integrate concepts within the theory of Knowledge Spaces. Ley, Kump and Albert (2010) have recently worked on combining expert judgement with Competetence-based Knowledge Spaces to enhance model building. Liu (2009) introduces a related technique to construct hierarchical knowledge structures from data.

As mentioned, the widely-used ALEKS tutor (`www.aleks.com`) is based on the Knowledge spaces theory. Although the commercial nature of ALEKS may imply that the latest developments and technical details are missing from the scientific literature, an instructive description can be found in (Falmagne et al., 2006).

Probabilistic assessment algorithms

The algorithms to assess the student's knowledge state in the knowledge spaces approach vary according to the version of the approach. In the simpler POKS framework, updating the probability that a student masters a given item can either follow a Naive Bayes approach as in (Desmarais et al., 2006), or slightly more sophisticated models such as a Tree Augmented Network (TAN) (Desmarais, 2011). In the more complex framework where knowledge structures are represented as AND/OR graphs, assessment of the student's knowledge state is modeled through a Markov Chain procedure, where the nodes are the potential student's knowledge state (there are 2^k possible states, where k is the number of items) and the probability that a student is in a given state is computed according to the Markov Chain model (Doignon and Falmagne, 1999). The ALEKS system relies on this procedure (Falmagne et al., 2006). A scheme similar to Falmagne et al.'s Markov Chain procedure was devised by Augustin et al. (2010) to assess skills with Compentence-based Knowledge Structures.

6. Assessing Learner Models

Learner models for use within adaptive systems should be validated, in order to guarantee that the model accurately assesses the construct(s)

that it is thought to assess. For static student models, this process can be as straightforward as assessing the power of a model to predict successes and failures from a subset of student response outcomes. Validation has also been extensively studied for static learner models such as IRT, for which statistics such as person-fit and item-fit are a standard part of most software packages (see Khalid, 2009, for a recent review). However, validation is more complex for dynamically changing learner models such as Cognitive Tutors and Constraint Based Modeling tutors. We review validation issues and techniques for dynamic student models in this section.

Assessments of changing student knowledge are typically validated with reference to two criteria. The first is the assessment's ability to predict future student performance within the learning system. The second is validation with external measures, such as post-tests of knowledge. In each case, actual student mastery is latent, and not directly measurable, but can be inferred via performance on other items designed to measure that same construct.

The first type of common validation, validating student performance with future performance within the learning system is conceptually similar to the use of factor analysis to measure static instruments (Cole, 1987). In the static case, any data on a given skill is equally relevant to predict any other data for that skill. In the dynamic case, by contrast, it is more relevant to see how the past predicts the future, than how the future predicts the past. More specifically, since the system's adaptive behavior at a given time will be based on assessment from evidence up to that time, the natural approach is to validate the assessment based on how the evidence up to a given time predicts performance on the student's next opportunity to demonstrate the skill. By aggregating evaluation of the assessment made at each possible prediction opportunity (e.g. each opportunity to practice the skill), it is possible to get an overall measure of how accurate the student model is.

As such, student performance between the first and n-th opportunity to practice a knowledge component, according the domain structure model, is inputted into the assessment system, and used to predict performance (correct or not correct) on the n+1-th opportunity to practice that knowledge component. The degree of correctness can be assessed using A', the probability that if the model is given two actions — one correct, the other incorrect — the model can accurately determine which is which. A' is also equivalent to the area under the ROC curve (Receiver Observer Characteristic), called AUC (Area Under the Curve)(Hanley and McNeil, 1982). An A' of 0.5 is equivalent to chance, and an A' of 1.0 represents perfect performance. A' has two useful properties — first, values of A' are comparable across models and data

sets — a model with A' of 0.54 is always better than a model with A' of 0.53. Second, A' values can be statistically compared to each other, or to chance, to establish the statistical significance of differences in A' (Fogarty et al., 2005). However, when conducting statistical significance tests on A' in learner behavior that occurs over time, it is important to take the non-independence of different observations of the same student into account, to avoid biasing in favor of statistical significance. A method for doing so is presented in (Baker et al., 2008a).

Other popular measures, such as kappa (Cohen, 1960), and accuracy, have significant disadvantages. Kappa, the degree to which a classifier is better than chance (Cohen, 1960), does not take uncertainty or probability into account, binarizing all probabilities before computation (e.g. 49% probability of correct is treated the same as 0% probability of correct, and 49% is treated as being the same distance from 51% as from 100%). As such, kappa has lower sensitivity to uncertainty, and can often give overly pessimistic estimates of an assessment's goodness, especially when used in a fail-soft or cost-sensitive fashion. In addition, kappa varies substantially based on changes in the proportion in the original data labels (DiEugenio and Glass, 2004), an issue when student performance is not close to 50%. Accuracy is a very simple measure, which divides agreements by all possible agreements. As such, accuracy does not take base rates into account. For instance, if a student gets 95% of answers correct, a system with an accuracy of 90% appears successful, but actually performs worse than a system which always guesses that the student will be correct.

An alternate metric is the Bayesian Information Criterion, BiC (Raftery, 1995). BiC takes both the degree of model fit and the model size into account, in order to account for the potential over-fit stemming from creating a model with too many parameters. BiC values can be compared for statistical significance between models in the same data set (Raftery, 1995).

Another way to control for over-fitting is to use k-fold cross-validation (Efrong and Gong, 1983) along with a metric such as A'. Within k-fold cross-validation, the data set is split into k groups ("folds"), and for each set of $k-1$ groups, a model is trained, and then tested on the k-th group. A variant, with subtly different properties but which is generally considered equally acceptable, is Leave-Out-One-Cross-Validation, where k is set equal to the number of data points. Cross-validation typically controls for over-fitting, at least at the level which cross-validation is conducted. As such, the level of cross-validation is key. For instance, one common error with learner data is to cross-validate at the level of individual actions. Cross-validating at the action level results in most students being represented in both training and test

folds. As such, cross-validating at this level validates that the model will predict new data from the same students. The goal of learner modeling is typically to develop models that will be accurate for new students who use the software after modeling efforts have completed, rather than generating models only usable for a specific group of students. It is possible to estimate a model's goodness for entirely new students by cross-validating at the student level, ensuring that each student in represented in only one data fold. This type of method is not currently explicitly supported in data mining/machine learning packages, but can be achieved using (for instance) Batch Cross-Validation in RapidMiner (Mierswa et al., 2006). Cross-validation can also be conducted at the level of units/lessons (a sub-segment of curriculum with a distinct user interface and set of skills), to validate that the model will be accurate for new curricular materials (cf. Baker et al., 2008), and at the level of schools, to validate that the model will generalize to new populations.

As mentioned earlier, another approach to validation is to validate assessments with external measures, such as post-tests of knowledge. The logic behind doing so is that learner behavior may be over-fit to the learning environment in some subtle fashion. For instance, the student may learn the cues associated with a correct answer, rather than learning a skill that will generalize outside of the learning environment, leading to overly high predictions of student skill. Corbett and Anderson (1995), for instance, found that Bayesian Knowledge Tracing in some cases predicted higher degrees of knowledge for slower learners than were borne out by post-test scores. By using an external post-test, it is possible to control for this possibility. In other cases, where direct assessment is not possible within the student's actions within the software (e.g. assessments of gaming the system), model predictions can be compared to outside assessments of behavior or self-report measures.

7. Other Key Areas in Student Modeling

7.1. Affect, Motivation, and Disengagement

In recent years, learner modeling has been extended to consider a broader range of aspects of the student. Researchers have begun to consider how to model key aspects of students' meta-cognition, motivation, and affect, towards providing adaptive scaffolding which can address individual differences in these areas. One key motivation for this development is the increasing awareness that it is difficult to address gaps in the student's knowledge, if the student is fundamentally unmotivated, and not taking the learning system seriously. Disengaged

students experience much lower learning, across forms of disengagement (cf. Baker, Corbett, Koedinger, & Wagner, 2004; Baker, 2007).

Affect has perhaps been the area which has received the greatest interest within learner modeling. A range of affective variables have been assessed within interactive learning environments, from relatively direct constructs such as emotional valence (positive or negative emotions), to Ekman's basic emotions (Ekman et al., 1987) such as anger, happiness, and fear, to the more cognitively complex OCC (Ortony, Clore, & Collins) model (Ortony et al., 1988) which includes states such as joy and shame, and recently to more cognitive-affective states that are more specific to the educational domain such as boredom, frustration, and uncertainty (cf. Litman & Forbes-Riley, 2004; D'Mello, Taylor, & Graesser, 2007). There has been considerable work towards deploying affect detection in educational software in the last decade. Conati's pioneering work in affect detection in educational software focused on a subset of the OCC model, and used a combination of physical sensors (electromyogram, heart rate, skin conductance) and distilled aspects of log files to detect student emotions as they played an educational game, PrimeClimb, in both laboratory settings and school settings (Conati et al., 2003; Conati and Maclaren, 2009a). Mota & Picard (2003) developed a model that could infer a student's interest from their posture. Litman and Forbes-Riley (2004) have developed software that can detect uncertainty from audio data. D'Mello and colleagues (2007) have developed models that use a combination of physical sensors (posture sensors, camera) and distilled aspects of log files to detect a range of cognitive-affective states thought to be particularly relevant for learning. Chaouachi and colleagues (2010) use affect detection based on EEG sensors to study student attention in educational software. In general, the common usage of physical sensors within this body of research has led to interesting findings and possibilities, but has limited the current applicability for large-scale deployment. Towards addressing this concern, D'Mello and colleagues (2008) have developed a variant of their affect detection software that uses no sensors. Arroyo et al. (2009) have created a relatively inexpensive suite of sensors (webcam, conductance bracelet, pressure mouse, posture analysis seat), and have succeeded in deploying detectors of confidence, frustration, excitement, and interest to entire classrooms at one time.

Motivation has been modeled by multiple efforts. de Vicente and Pain (2002) developed a model that could detect several aspects of motivation, including desire for control, challenge, and independence. Similarly, Rebolledo-Mendez, du Boulay, and Luckin (2006) modeled effort, confidence, and independence within a vygotskyan intelligent

tutor. Conati and Maclaren (2009) modeled learner goals within Prime-Climb, by correlating student responses on goal-orientation questionnaires with their interactive behaviors within the game. Conati and Maclaren's framework integrated prediction of student goals with prediction of student affect, and even prediction of student personality characteristics, towards developing complete learner models within their game.

Other research has focused on modeling behaviors associated with disengagement. For example, automated detectors of gaming the system (attempting to succeed in an educational task by systematically taking advantage of properties and regularities in the system used to complete that task, rather than by thinking through the material) have been developed for several learning systems (Baker et al., 2008b; Baker and de Carvalho, 2008; Baker et al., 2010; Beal et al., 2006; Beck, 2005; Johns and Woolf, 2006; Muldner et al., 2010; Walonoski and Heffernan, 2006), and have been used as the basis of automated interventions that reduced gaming and improved learning (Arroyo et al., 2007; Baker et al., 2006). Detectors of off-task behavior (Baker, 2007; Cetintas et al., 2009) have also been developed and validated.

7.2. Meta-Cognition and Self-Regulated Learning

Meta-cognition, "active monitoring and consequent regulation and orchestration of cognitive processes to achieve cognitive goals" (Hacker, 1999), has received interest in learner modeling in recent years. Aleven et al. (2006) developed a meta-cognitive model which unified several different constructs related to help-seeking and help avoidance behavior in a single model. This model, developed for Cognitive Tutors, was embedded into a Cognitive Tutor and used to provide meta-cognitive feedback (Roll et al., 2009). Biswas et al. (2010) developed a self-regulated learning model that can identify a range of behaviors in the space of self-regulated learning/meta-cognition, including monitoring through explanation, self-assessment, and setting learning goals. Montalvo et al. (2010) presented models that could detect two forms of planning behavior in Science Microworlds. Shih et al. (2008) developed a model that can infer whether a student is self-explaining within an intelligent tutor.

7.3. Open Learner Modeling

An open learner model "makes a machine's representation of the learner available as an important means of support for learning" (Bull and Kay, 2010). Though open learner models are outside the main scope of this review, they are an important ongoing area of learner modeling, often

applying to the context of skill models, while interacting in interesting fashions with meta-cognitive modeling. Bull and Kay (2007) have identified several areas of potential contribution for open learner models, including (but not limited to): promoting meta-cognitive activities, encouraging learner independence, promoting collaboration and competition, and increasing learner knowledge about the learning system and trust in the learning system. One major ongoing question in open learner modeling is how much learner control should be allowed, with considerable ongoing research in how to best support and leverage negotiation and persuasion between learners and open learner models (Bull and Pain, 1995; Dimitrova, 2003; Mabbott and Bull, 2006), towards improving model accuracy, effectiveness, and student trust in the learner model. While the systems we have discussed in this paper do not open their learner models to the degree proposed by Bull and colleagues, many of these systems allow students to see the system's assessments of their learning, for instance through "skill bars" (Koedinger et al., 1997) indicating the probability that the student knows each skill.

7.4. Group and Collaborative Learner Modeling

Although much of learner modeling has taken place in the context of individual learning, a great deal of learning, both online and off-line, takes place in the context of groups and collaborations. Kay et al. (2006) and Perera et al. (2009) have developed models of teamwork strategies and individual behaviors in group learning, studying which behaviors of teams and individuals lead to successful group-work in educational settings. Walker and colleagues (2010) have developed models of helping behaviors in computer-mediated peer tutoring, towards improving peer tutor help. Rosé and colleagues have developed automated detection of insults and other off-topic behaviors (Kumar et al., 2007; Prata et al., 2009). Vassileva and colleagues have developed systems for identifying which student in a population is best suited for supporting the learning of another, struggling student(Vassileva et al., 2003). Bull and Broady (1997) have developed models of how student models can be shared to create in-the-moment peer tutoring episodes.

7.5. Long-Term Learner Modeling

One limitation present in many widely used learing environments is the isolated nature of their learner models. Incredibly rich representations of students' knowledge are created, refined, and then discarded at the end of the school year, although this information could be extremely useful in future learning (both immediately, such as in the next school year, and in life-long learning), and also in other classes

occurring at the same time. As education increasingly moves online, a student may encounter overlapping content multiple times during their learner trajectory — shared learner models could prevent boring and time-consuming redundancy (correspondingly, these models could also enable more multi-pronged response to a student difficulty or misconception).

Coordinating the sharing and interoperability of learner models across learning environments is an area which has been the subject of considerable research in recent years. Several repositories of open educational resources have been created (Hatala et al., 2004; Alrifai et al., 2006), as well as standards for interoperability of learner models and learning environments (Friesen, 2005). Models have been articulated for how to represent learner information in an interoperable fashion (Aroyo et al., 2006), for how to import learner models into a new learning environment (Dolog and Schäfer, 2005), and for how to exchange information between learner models using ontologies (Brusilovsky et al., 2005). A compelling "ecological" vision for how information of general value can be captured, shared, and used among learning environments is presented by McCalla et al. (2004). A vision for how this type of support could eventually scaffold life-long learning is suggested by (Koper et al., 2005). The potential of this approach is supported by recent empirical work by Pardos and Heffernan (2010), in improving individualization through developing and utilizing student-level assessments when the student encounters new material .

However, despite the increasing sophistication of practice and theory in this area, sharing of data between learner environments has not yet emerged into the most widely-used learning environments, such as those detailed within this paper. One key challenge is that ontologies need to be aligned across learning systems; mis-alignments may actually cause students to miss needed content. A second issue is that using information from a different learning environment presents some risk to a content developer — if the other system's model is inaccurate, then the system using that information may act incorrectly through no fault of its own. Privacy issues also may limit the degree of information-sharing. Nonetheless, as these obstacles are addressed, it can be anticipated that sharing of information between learner models will become more widespread, leading to improvements in education.

8. Conclusion

In the last decade, intelligent learning environments based on sophisticated learner models have emerged into much wider use than seen

previously, with the most widely-used systems being used by tens or hundreds of thousands of learners a year. These systems embed learner models that effectively deal with uncertainty and partial evidence. Recent advances increasingly allow them to adapt not just based on which skills students know, but also based on assessments of complex meta-cognitive, motivational, and affective constructs. These environments increasingly can adapt not just to a single learner but can also support collaborative and group learning. As well as adapting, these systems often open their learner models to at least some degree, supporting students in learning more about their own learning progress.

One direction that is becoming of increasing prominence in learner modeling is the use of educational data mining techniques. As discussed in this article, data mining methods have supported the emergence of both automated domain model search and models of meta-cognition, motivation, and affect. As more and more learner data becomes available, and methods for exploiting that data improve, the potential is present for better and better learner models.

Learner models are often developed using a combination of methods, including data mining as well as knowledge engineering carried out by domain experts; increasingly, effective learner models are therefore developed by inter-disciplinary teams. One trend is that these teams often carry out their tasks separately, even though their efforts are eventually integrated into a single learner model. Hence, one group may use knowledge-engineering to develop a domain model, and then another group may use data mining to parameterize a Knowledge Tracing model that operates on that domain model.

Among the difficult challenges that remain is the sharing of learner models across learning systems, as discussed earlier. However, in the long-term, this trend may lead to a more integrated and effective educational experience for students, across their life-time of learning.

In the long-term, as the field gets better at developing, refining, and exploiting sophisticated multi-dimensional models of learners, there is improved potential for tailoring each student's learning experiences to their educational needs. At the current rate of progress, we look forward to transformative progress in learner modeling by the time of the 30th anniversary issue of UMUAI!

References

Aleven, V., B. M. McLaren, I. Roll, and K. R. Koedinger: 2006, 'Toward Meta-cognitive Tutoring: A Model of Help Seeking with a Cognitive Tutor'. *I. J. Artificial Intelligence in Education* **16**(2), 101–128.

Almond, R. G., L. Dibello, F. Jenkins, D. Senturk, R. Mislevy, L. Steinberg, and D. Yan: 2001, 'Models for conditional probability tables in educational assessment'. In: *Proceedings of the 2001 Conference on AI and Statistics.*

Almond, R. G., L. V. DiBello, B. Moulder, and J.-D. Zapata-Rivera: 2007, 'Modeling Diagnostic Assessments with Bayesian Networks'. *Journal of Educational Measurement* **44**, 341–359. 10.1111/j.1745-3984.2007.00043.x.

Almond, R. G. and R. J. Mislevy: 1999, 'Graphical models and computerized adaptive testing'. *Applied Psychological Measurement* **23**(3), 223–237.

Alrifai, M., P. Dolog, and W. Nejdl: 2006, 'Learner Profile Management for Collaborative Adaptive eLearning Application'. In: *APS'2006: Joint International Workshop on Adaptivity, Personalisation and the Semantic Web at the 17th ACM Hypertext'06 conference.*

Aroyo, L., P. Dolog, G.-J. Houben, M. Kravcik, A. Naeve, M. Nilsson, and F. Wild: 2006, 'Interoperability in Personalized Adaptive Learning'. *Educational Technology & Society* **9**(2), 4–18.

Arroyo, I., D. G. Cooper, W. Burleson, B. P. Woolf, K. Muldner, and R. Christopherson: 2009, 'Emotion Sensors Go To School'. In: V. Dimitrova, R. Mizoguchi, B. du Boulay, and A. C. Graesser (eds.): *Proceedings of the 14th International Conference on Artificial Intelligence in Education, AIED 2009, July 6-10, 2009, Brighton, UK*. pp. 17–24.

Arroyo, I., K. Ferguson, J. Johns, T. Dragon, H. Meheranian, D. Fisher, A. G. Barto, S. Mahadevan, and B. P. Woolf: 2007, 'Repairing Disengagement With Non-Invasive Interventions'. In: R. Luckin, K. R. Koedinger, and J. E. Greer (eds.): *Proceedings of the 13th International Conference on Artificial Intelligence in Education, AIED 2007, July 9-13, 2007, Los Angeles, California, USA*. pp. 195–202.

Augustin, T., C. Hockemeyer, M. Kickmeier-Rust, and D. Albert: 2010, 'Individualized Skill Assessment in Digital Learning Games: Basic Definitions and Mathematical Formalism'. *IEEE Transactions on Learning Technologies* **99**(PrePrints).

Baker, F. B. and S.-H. Kim: 2004, *Item Response Theory, Parameter Estimation Techniques (2nd ed.)*. New York, NY: Marcel Dekker Inc.

Baker, R. and A. Corbett: 2008, 'More Accurate Student Modeling through Contextual Estimation of Slip and Guess Probabilities in Bayesian Knowledge Tracing'. In: *Proceedings of Ninth Intelligent Tutoring System Conference (ITS2008)*. pp. 406–415. 10.1007/978-3-540-69132-7_44.

Baker, R. S., A. T. Corbett, K. R. Koedinger, and A. Z. Wagner: 2004, 'Off-Task Behavior in the Cognitive Tutor Classroom: When Students "Game The System"'. In: *Proceedings of ACM CHI 2004: Computer-Human Interaction*. pp. 383–390.

Baker, R. S. J.: 2007, 'Modeling and understanding students' off-task behavior in intelligent tutoring systems'. In: M. B. Rosson and D. J. Gilmore (eds.): *Proceedings of the 2007 Conference on Human Factors in Computing Systems, CHI 2007, San Jose, California, USA, April 28 - May 3, 2007*. pp. 1059–1068.

Baker, R. S. J., A. T. Corbett, and V. Aleven: 2008a, 'Improving Contextual Models of Guessing and Slipping with a Trucated Training Set'. In: R. S. J. Baker, T. Barnes, and J. E. Beck (eds.): *Proceedings of EDM 2008, The 1st International Conference on Educational Data Mining*. pp. 67–76.

Baker, R. S. J., A. T. Corbett, K. R. Koedinger, S. Evenson, I. Roll, A. Z. Wagner, M. Naim, J. Raspat, D. J. Baker, and J. E. Beck: 2006, 'Adapting to When Students Game an Intelligent Tutoring System'. In: M. Ikeda, K. D. Ashley, and T.-W. Chan (eds.): *Intelligent Tutoring Systems, 8th International Conference,*

ITS 2006, Jhongli, Taiwan, June 26-30, 2006, Proceedings, Vol. 4053 of *Lecture Notes in Computer Science*. pp. 392–401.

Baker, R. S. J., A. T. Corbett, I. Roll, and K. R. Koedinger: 2008b, 'Developing a generalizable detector of when students game the system'. *User Model. User-Adapt. Interact* **18**(3), 287–314.

Baker, R. S. J. and A. M. J. B. de Carvalho: 2008, 'Labeling Student Behavior Faster and More Precisely with Text Replays'. In: *Proceedings of the 1st International Conference on Educational Data Mining, EDM 2008*. pp. 38–47.

Baker, R. S. J., A. Mitrovic, and M. Mathews: 2010, 'Detecting Gaming the System in Constraint-Based Tutors'. In: *User Modeling, Adaptation, and Personalization, 18th International Conference, UMAP 2010, Big Island, HI, USA, June 20-24, 2010. Proceedings*. pp. 267–278.

Beal, C. R., L. Qu, and H. Lee: 2006, 'Classifying Learner Engagement through Integration of Multiple Data Sources'. In: *Prooceedings of AAAI06*.

Beck, J. E.: 2005, 'Engagement tracing: using response times to model student disengagement'. In: C.-K. Looi, G. I. McCalla, B. Bredeweg, and J. Breuker (eds.): *Proceedings of the 12th International Conference on Artificial Intelligence in Education, AIED 2005, July 18-22, 2005, Amsterdam, The Netherlands*. pp. 88–95.

Beck, J. E., K. min Chang, J. Mostow, and A. T. Corbett: 2008, 'Does Help Help? Introducing the Bayesian Evaluation and Assessment Methodology'. In: *Proceedings of Intelligent Tutoring Systems,ITS 2008, Montreal, Canada, June 23-27, 2008*. pp. 383–394.

Biswas, G., H. Jeong, J. Kinnebrew, B. Sulcer, , and R. Roscoe: 2010, 'Measuring Self-regulated Learning Skills through Social Interactions in a Teachable Agent Environment.'. *Research and Practice in Technology-Enhanced Learning* **5**(2), 123–152.

Bloom, B. S.: 1984, 'The 2 sigma problem: The search for methods of group instruction as effective as one-to-one tutoring'. *Educational Researcher* **13**(4), 4–16.

Briggs, D. C. and M. Wilson: 2003, 'An introduction to multidimensional measurement using Rasch models'. *Journal of Applied Measurement* pp. 87–100.

Brusilovsky, P.: 2001, 'Adaptive Hypermedia'. *User Modeling and User-Adapted Interaction: The Journal of Personalization Research* **11**(1-2), 87–110.

Brusilovsky, P.: 2003, 'Adaptive and intelligent web-based educational systems'. *International Journal of Artificial Intelligence in Education* **13**, 156–169.

Brusilovsky, P., S. Sosnovsky, and M. Yudelson: 2005, 'Ontology-based Framework for User Model Interoperability in Distributed Learning Environments'. In: *World Conference on E-Learning, E-Learn 2005*. pp. 2851–2855.

Brusilovsky, P. and J. Vassileva: 2003, 'Course Sequencing Techniques for Large-Scale Webbased Education'.

Bull, S., , and E. Broady: 1997, 'Spontaneous peer tutoring from sharing student models.'. In: *Proceedings of the 8th World Conference on Artificial Intelligence in Education (AI-ED97)*. pp. 143–150.

Bull, S. and J. Kay: 2007, 'Student Models that Invite the Learner In: The SMILI Open Learner Modelling Framework.'. *International Journal of Artificial Intelligence in Education* **17**(2), 89–120.

Bull, S. and J. Kay: 2010, 'Open Learner Models'. In: R. Nkambou and R. Bourdeau, J. Mizoguchi (eds.): *Advances in Intelligent Tutoring Systems*. pp. 301–322.

Bull, S. and H. Pain: 1995, 'Did I say what I think I said, and do you agree with me?: Inspecting and Questioning the Student Model.'. In: J. Greer (ed.): *Proceedings of World Conference on Artificial Intelligence and Education*. pp. 501–508.

Carmona, C., E. Millán, J.-L. P. de-la Cruz, M. Trella, and R. Conejo: 2005, 'Introducing Prerequisite Relations in a Multi-layered Bayesian Student Model'. In: L. Ardissono, P. Brna, and A. Mitrovic (eds.): *User Modeling 2005, 10th International Conference, UM 2005*. pp. 347–356.

Cen, H., K. R. Koedinger, and B. Junker: 2006, 'Learning Factors Analysis — A General Method for Cognitive Model Evaluation and Improvement'. In: *Intelligent Tutoring Systems, 8th International Conference, ITS 2006, Jhongli, Taiwan, June 26-30, 2006, Proceedings*. pp. 164–175.

Cetintas, S., L. Si, Y. P. Xin, C. Hord, and D. Zhang: 2009, 'Learning to Identify Students' Off-Task Behavior in Intelligent Tutoring Systems'. In: V. Dimitrova, R. Mizoguchi, B. du Boulay, and A. C. Graesser (eds.): *Proceedings of the 14th International Conference on Artificial Intelligence in Education, AIED 2009, July 6-10, 2009, Brighton, UK*. pp. 701–703.

Chaouachi, M. and C. Frasson: 2010, 'Exploring the Relationship between Learner EEG Mental Engagement and Affect'. In: V. Aleven, J. Kay, and J. Mostow (eds.): *Intelligent Tutoring Systems, 10th International Conference, ITS 2010, Pittsburgh, PA, USA, June 14-18, 2010, Proceedings, Part II*. pp. 291–293.

Cohen, J. A.: 1960, 'A coefficient of agreement for nominal scales'. *Educational and Psychological Measurement* **20**(1), 37–46.

Cohen, P. A., J. A. Kulik, and C. L. C. Kulik: 1982, 'Educational outcomes of tutoring: A meta-analysis of findings'. *American Educational Research Journal* **19**, 237–248.

Cole, D. A.: 1987, 'tility of confirmatory factor analysis in test validation research'. *Journal of Consulting and Clinical Psychology* **55**(4), 584–594.

Conati, C., R. Chabbal, and H. Maclaren: 2003, 'A Study on Using Biometric Sensors for Detecting User Emotions in Educational Games'. In: *Proceedings of the Workshop "Assessing and Adapting to UserAttitude and Affects: Why, When and How?"*.

Conati, C., A. Gertner, and K. VanLehn: 2002, 'Using Bayesian Networks to Manage Uncertainty in Student Modeling'. *User Modeling and User-Adapted Interaction* **12**(4), 371–417.

Conati, C. and H. Maclaren: 2009a, 'Empirically building and evaluating a probabilistic model of user affect'. *User Model. User-Adapt. Interact* **19**(3), 267–303.

Conati, C. and H. Maclaren: 2009b, 'Empirically building and evaluating a probabilistic model of user affect'. *User Model. User-Adapt. Interact* **19**(3), 267–303.

Conejo, R., E. Guzman, E. Millán, M. Trella, J. L. Pérez-de-la Cruz, and A. Rios: 2004, 'SIETTE: A Web-Based Tool for Adaptive Teaching'. *International Journal of Artificial Intelligence in Education* **14**, 29–61.

Corbett, A. T. and J. R. Anderson: 1995, 'Knowledge Tracing: Modeling the Acquisition of Procedural Knowledge'. *User Modeling and User-Adapted Interaction* **4**(4), 253–278.

Cumming, G. and A. Mcdougall: 2000, 'Mainstreaming AIED into Education?'. *International Journal of Artificial Intelligence in Education (IJAIED)* **11**, 197–207.

de Vicente, A. and H. Pain: 2002, 'Informing the Detection of the Students' Motivational State: An Empirical Study'. *Lecture Notes in Computer Science* **2363**, 933–943.

Desmarais, M. C.: 2011, 'Performance Comparison of Item-to-Item Skills Models with the IRT Single Latent Trait Model'. In: *Proceedings of UMAP2011 Conference*.

Desmarais, M. C., A. Maluf, and J. Liu: 1996, 'User-Expertise Modeling with Empirically Derived Probabilistic Implication Networks'. *User Modeling and User-Adapted Interaction* **5**(3-4), 283–315.

Desmarais, M. C., P. Meshkinfam, and M. Gagnon: 2006, 'Learned student models with item to item knowledge structures'. *User Modeling and User-Adapted Interaction* **16**(5), 403–434.

DiEugenio, B. and M. Glass: 2004, 'The Kappa Statistic: A Second Look'. *Computational Linguistics* **30**(1), 95–101.

Dimitrova, V.: 2003, 'StyLE-OLM: Interactive Open Learner Modelling'. *International Journal of Artificial Intelligence in Education* **13**(1), 35–78.

D'Mello, S. K., S. D. Craig, A. M. Witherspoon, B. McDaniel, and A. C. Graesser: 2008, 'Automatic detection of learner's affect from conversational cues'. *User Model. User-Adapt. Interact* **18**(1-2), 45–80.

D'Mello, S. K., R. S. Taylor, and A. Graesser: 2007, 'Affective Trajectories during Complex Learning'. In: *Proceedings of the 29th Annual Meeting of the Cognitive Science Society, Austin, TX*. pp. 203–208.

Doignon, J.-P. and J.-C. Falmagne: 1985, 'Spaces for the assessment of knowledge'. *International Journal of Man-Machine Studies* **23**, 175–196.

Doignon, J.-P. and J.-C. Falmagne: 1999, *Knowledge Spaces*. Berlin: Springer-Verlag.

Dolog, P. and M. Schäfer: 2005, 'A Framework for Browsing, Manipulating and Maintaining Interoperable Learner Profiles.'. In: *Proceedings of UM'2005*. pp. 397–401.

Dütsch, I. and G. Gediga: 1995, 'Skills and knowledge structures'. *British Journal of Mathematical and Stastical Psychology* **48**, 9–27.

Efrong, B. and G. Gong: 1983, 'A leisurely look at the bootstrap, the jackknife, and cross-validation'. *American Statistician* **37**, 36–48.

Ekman, P., W. V. Friesen, M. O'Sullivan, A. Chan, I. Diacoyanni-Tarlatzi, K. Heider, R. Krause, W. A. Lecompte, T. Pitcairn, P. E. Picci-Bitti, K. Sherer, M. Tomita, and A. Tzavaras: 1987, 'Universal and Cultural Differences in the Judgment of Facial Expression and Emotiosn'. *Journal of Personality and Social Psychology* **53**, 712–717.

Falmagne, J.-C., E. Cosyn, J.-P. Doignon, and N. Thiéry: 2006, 'The Assessment of Knowledge, in Theory and in Practice'. In: R. Missaoui and J. Schmid (eds.): *ICFCA*, Vol. 3874 of *Lecture Notes in Computer Science*. pp. 61–79.

Feng, M., N. T. Heffernan, and K. R. Koedinger: 2006, 'Addressing the testing challenge with a web-based e-assessment system that tutors as it assesses'. In: L. Carr, D. D. Roure, A. Iyengar, C. A. Goble, and M. Dahlin (eds.): *Proceedings of the 15th international conference on World Wide Web, WWW 2006, Edinburgh, Scotland, UK, May 23-26, 2006*. pp. 307–316.

Fogarty, J., R. S. Baker, and S. E. Hudson: 2005, 'Case studies in the use of ROC curve analysis for sensor-based estimates in human computer interaction'. In: K. Inkpen and M. van de Panne (eds.): *Proceedings of the Graphics Interface 2005 Conference, May 9-11, 2005, Victoria, British Columbia, Canada*. pp. 129–136.

Forbes-Riley, K. and D. J. Litman: 2004, 'Predicting Emotion in Spoken Dialogue from Multiple Knowledge Sources'. In: *HLT-NAACL*. pp. 201–208.

Friesen, N.: 2005, 'Interoperability and learning objects: An overview of e-learning standardization.'. *Interdisciplinary Journal of Knowledge and Learning Objects* **1**, 23–31.

Gong, Y., J. E. Beck, and N. T. Heffernan: 2010, 'Using multiple Dirichlet distributions to improve parameter plausibility'. In: R. S. J. de Baker, A. Merceron, and P. I. P. Jr (eds.): *Educational Data Mining 2010 (EDM2010), Pittsburgh, PA, USA, June 11-13, 2010. Proceedings*. pp. 61–70.

Greer, J. E. and G. I. McCalla: 1989, 'A Computational Framework for Granularity and its Application to Educational Diagnosis'. In: *IJCAI*. pp. 477–482.

Guzmán, E., R. Conejo, and J.-L. Pérez-de-la Cruz: 2007, 'Adaptive testing for hierarchical student models'. *User Modeling and User-Adapted Interaction* **17**, 119–157.

Hacker, D. J.: 1999, 'Definitions and Empirical Foundations'. In: D. J. Hacker, J. Dunlosky, and A. C. Graesser (eds.): *Metacognition in Educational Theory and Practice*. pp. 1–24.

Haertel, E. H.: 1989, 'Using Restricted Latent Class Models to Map the Skill Structure of Achievement Items'. *Journal of Educational Measurement* **26**, 301–321. 10.1111/j.1745-3984.1989.tb00336.x.

Hanley, J. A. and B. J. McNeil: 1982, 'The Meaning and Use of the Area Under a Receiver Operating Characteristic (ROC) Curve'. *Radiology*.

Hastie, T., R. Tibshirani, and J. H. Friedman: 2001, *The Elements of Statistical Learning*. Springer.

Hatala, M., G. Richards, T. Eap, , and J. Willms: 2004, 'Interoperability of Learning Object Repositories and Services: Standards, Implementations and Lessons Learned'. In: *The Proceedings of the 13th World Wide Web Conference*.

Heller, J., C. Steiner, C. Hockemeyer, and D. Albert: 2006, 'Competence–Based Knowledge Structures for Personalised Learning'. *International Journal on E–Learning* **5**(1), 75–88.

Hockemeyer, C., T. Held, and D. Albert: 1997, 'RATH - A Relational Adaptive Tutoring Hypertext WWW-Environment Based on Knowledge Space Theory'.

Jameson, A.: 1995, 'Numerical Uncertainty Management in User and Student Modeling: An Overview of Systems and Issues'. *User Modeling and User-Adapted Interaction* **5**(3-4), 193–251.

Johns, J. and B. P. Woolf: 2006, 'A Dynamic Mixture Model to Detect Student Motivation and Proficiency'. In: *Proceedings of AAAI2006*.

Junker, B. and K. Sijtsma: 2001, 'Cognitive Assessment Models with Few Assumptions, and Connections With Nonparametric Item Response Theory'. *Applied Psychological Measurement* **25**(3), 258–272.

Kay, J., N. Maisonneuve, K. Yacef, and P. Reimann: 2006, 'The Big Five and Visualisations of Team Work Activity'. In: *Proceedings of Intelligent Tutoring Systems (ITS06)*. pp. 197–206.

Khalid, M. N.: 2009, 'IRT model fit from different perspectives'. Ph.D. thesis, University of Twente.

Koedinger, K. R., J. R. Anderson, W. H. Hadley, and M. A. Mark: 1997, 'Intelligent tutoring goes to school in the big city'. *International Journal of Artificial Intelligence in Education* **8**, 30–43.

Koedinger, K. R., A. T. Corbett, and C. Perfetti: 2011, 'The Knowledge-Learning-Instruction (KLI) Framework: Toward Bridging the Science-Practice Chasm to Enhance Robust Student Learning'. (under review).

Koper, E. J. R., B. Giesbers, P. Van Rosmalen, P. Sloep, J. Van Bruggen, C. Tattersall, H. Vogten, and F. Brouns: 2005, 'A design model for lifelong learning networks'. *Interactive Learning Environments* **1–2**, 71–92.

Kumar, R., C. P. Rosé, Y. C. Wang, M. Joshi, , and A. Robinson: 2007, 'Tutorial dialogue as adaptive collaborative learning support'. In: *Proceedings of the 13th International Conference on Artificial Intelligence in Education (AIED 2007)*.

Lajoie, S.: 2005, 'Extending the scaffolding metaphor'. *Instructional Science* **33**, 541–557.

Lepper, M. R., M. Woolverton, D. L. Mumme, and J. L. Gurtner: 1991, 'Motivational techniques of expert human tutors: lessons for the design of computer-based tutors'. In: S. P. Lajoie and S. J. Derry (eds.): *Computers as cognitive tools. Hillsdale*.

Ley, T., B. Kump, and D. Albert: 2010, 'A methodology for eliciting, modelling, and evaluating expert knowledge for an adaptive work-integrated learning system'. *Int. J. Hum.-Comput. Stud.* **68**(4), 185–208.

Liu, C.-L.: 2009, 'A Simulation-Based Experience in Learning Structures of Bayesian Networks to Represent How Students Learn Composite Concepts'. *Behaviormetrika* **36**(1), 1–25.

Mabbott, A. and S. Bull: 2006, 'Student Preferences for Editing, Persuading and Negotiating the Open Learner Model'. In: *Proceedings of the International Conference on Intelligent Tutoring Systems (ITS 2006)*. pp. 481–490.

Mayo, M. and A. Mitrovic: 2001, 'Optimising ITS Behaviour with Bayesian Networks and Decision Theory'. *International Journal of Artificial Intelligence in Education* **12**, 124–153.

McCalla, G.: 2004, 'The ecological approach to the design of e-learning environments: purpose-based capture and use of information about learners'. *Journal of Interactive Media in Education* **7**.

McCalla, G., J. Greer, B. Barrie, and P. Pospisil: 1992, 'Granularity hierarchies'. *Computers & Mathematics with Applications* **23**(2–5), 363–375.

McCalla, G. I., D. R. Peachey, and B. Ward: 1982, 'An architecture for the design of large-scale intelligent teaching systems'. In: N. Cercone and G. McCalla (eds.): *Prodeedings of the 4th National Conference of the CSCSI*. pp. 85–91.

Mierswa, I., M. Wurst, R. Klinkenberg, M. Scholz, and T. Euler: 2006, 'YALE: Rapid Prototyping for Complex Data Mining Tasks'. In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2006)*. pp. 935–940.

Millán, E. and J. L. Pérez-de-la-Cruz: 2002, 'A Bayesian Diagnostic Algorithm for Student Modeling and its Evaluation'. *User Modeling and User-Adapted Interaction* **12**(2–3), 281–330,.

Mislevy, R. J., R. G. Almond, D. Yan, and L. S. Steinberg: 1999, 'Bayes Nets in Educational Assessment: Where the Numbers Come From'. In: K. B. Laskey and H. Prade (eds.): *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence (UAI-99)*. S. F., Cal., pp. 437–446.

Mislevy, R. J. and D. Gitomer: 1995, 'The Role of Probability-Based Inference in an Intelligent Tutoring System'. *User Modeling and User-Adapted Interaction* **42**(5), 253–282.

Mitrovic, A.: 2011, 'Fifteen years of Constraint-Based Tutors: What we have achived and where we are going'. *User Modeling and User-Adapted Interaction*.

Mitrovic, A., K. R. Koedinger, and B. Martin: 2003, 'A Comparative Analysis of Cognitive Tutoring and Constraint-Based Modeling'. In: *User Modeling 2003,*

9th International Conference, UM 2003, Johnstown, PA, USA, June 22-26, 2003. pp. 313–322.

Montalvo, O., R. S. J. Baker, M. A. Sao Pedro, A. Nakama, and J. D. Gobert: 2010, 'Identifying Student' Inquiry Planning Using Machine Learning'. In: *Proceedings of the 3rd International Conference on Educational Data Mining.* pp. 141–150.

Mota, S. and R. Picard: 2003, 'Automated Posture Analysis for Detecting Learner's Interest Level'. In: *Proceedings of the CVPR Workshop Computer Vision and Pattern Recognition for Human Computer Interaction.*

Muldner, K., W. Burleson, B. van de Sande, and K. VanLehn: 2010, 'An Analysis of Gaming Behaviors in an Intelligent Tutoring System'. In: V. Aleven, J. Kay, and J. Mostow (eds.): *Intelligent Tutoring Systems, 10th International Conference, ITS 2010, Pittsburgh, PA, USA, June 14-18, 2010, Proceedings, Part I*, Vol. 6094 of *Lecture Notes in Computer Science.* pp. 184–193.

Neapolitan, R. E.: 2004, *Learning Bayesian Networks.* New Jersey: Prentice Hall.

Nkambou, R., J. Bourdeau, and R. Mizoguchi (eds.): 2010, *Advances in Intelligent Tutoring Systems.* Springer.

Ortony, A., G. L. Clore, and A. Collins: 1988, *The cognitive structure of emotions.* Cambridge University Press.

Pardos, Z. and N. Heffernan: 2011, 'KT-IDEM: Introducing Item Difficulty to the Knowledge Tracing Model'. In: *User Modeling, Adaptation, and Personalization, 19th International Conference, UMAP 2011, Gerona, Spain, July 11–15, 2011. Proceedings.* p. (to appear).

Pardos, Z. A. and N. Heffernan: 2010a, 'Navigation the parameter space of Bayesian Knowledge Tracing models: Visualizations of the convergence of the Expectation Maximization algorithm'. In: *Proceedings of the 3rd Educational Data Mining Conference 2010.* pp. 161–170.

Pardos, Z. A. and N. T. Heffernan: 2010b, 'Modeling Individualization in a Bayesian Networks Implementation of Knowledge Tracing'. In: *Proceedings of the 18th International Conference on User Modeling Adaptation and Personalization.*

Pavlik, P. I., H. Cen, and K. R. Koedinger: 2009a, 'Learning Factors Transfer Analysis: Using Learning Curve Analysis to Automatically Generate Domain Models'. In: T. Barnes, M. C. Desmarais, C. Romero, and S. Ventura (eds.): *Educational Data Mining - EDM 2009, Cordoba, Spain, July 1-3, 2009. Proceedings of the 2nd International Conference on Educational Data Mining.* pp. 121–130.

Pavlik, P. I., H. Cen, and K. R. Koedinger: 2009b, 'Performance Factors Analysis –A New Alternative to Knowledge Tracing'. In: *Proceeding of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling.* Amsterdam, The Netherlands, The Netherlands, pp. 531–538.

Peachey, D. R. and G. I. McCalla: 1986, 'Using planning techniques in intelligent tutoring systems'. *International Journal of Man-Machine Studies* **24**(1), 77–98.

Perera, D., J. Kay, I. Koprinska, K. Yacef, , and O. R. Zaïane: 2009, 'Clustering and Sequential Pattern Mining of Online Collaborative Learning Data.'. *IEEE Transactions on Knowledge and Data Engineering* **21**(6), 759–772.

Prata, D. N., R. S. J. d. Baker, E. Costa, C. P. Rosé, Y. Cui, and A. M. J. B. de Carvalho: 2009, 'Detecting and Understanding the Impact of Cognitive and Interpersonal Conflict in Computer Supported Collaborative Learning Environments'. In: *Proceedings of the 2nd International Conference on Educational Data Mining.* pp. 131–140.

Raftery, A. E.: 1995, 'Bayesian model selection in social research (with discussion)'. *Sociological Methodology* **25**, 111–193.

Rebolledo-Mendez, G., B. du Boulay, , and R. Luckin: 2006, 'Motivating the learner: an empirical evaluation'. In: *8th International Conference on Intelligent Tutoring Systems (ITS2006*. pp. 545–554.

Reckase, M. D. and R. L. McKinley: 1991, 'The Discriminating Power of Items That Measure More Than One Dimension'. *Applied Psychological Measurement* **15**(4), 361–373.

Reye, J.: 2004, 'Student Modelling based on Belief Networks'. *International Journal of Artificial Intelligence in Education* **14**, 63–96.

Roll, I., V. Aleven, B. M. Mclaren, and K. R. Koedinger: 2009, 'Can Help Seeking Be Tutored? Searching for the Secret Sauce of Metacognitive Tutoring'. In: *Artificial Intelligence in Education (AIED 2007)*. pp. 203–210.

Roussos, L. A., J. L. Templin, and R. A. Henson: 2007, 'Skills Diagnosis Using IRT-Based Latent Class Models'. *Journal of Educational Measurement*.

Self, J.: 1988, 'Bypassing the intractable problem of student modelling'. In: *Proceedings of Intelligent Tutoring Systems, ITS'88*. pp. 18–24.

Shih, B., K. R. Koedinger, and R. Scheines: 2008, 'A response time model for bottom-out hints as worked examples'. In: *Proceedings of the First International Conference on Educational Data Mining*. pp. 117–126.

Stout, W.: 2007, 'Skills Diagnosis Using IRT-Based Continuous Latent Trait Models'. *Journal of Educational Measurement* **44**, 313–324. 10.1111/j.1745-3984.2007.00041.x.

Tatsuoka, K. K.: 1983, 'Rule space: An approach for dealing with misconceptions based on item response theory'. *Journal of Educational Measurement* **20**, 345–354.

Vanlehn, K., C. Lynch, K. Schulze, J. A. Shapiro, R. Shelby, L. Taylor, D. Treacy, A. Weinstein, and M. Wintersgill: 2005, 'The Andes Physics Tutoring System: Lessons Learned'. *Int. J. Artif. Intell. Ed.* **15**(3), 147–204.

VanLehn, K., Z. Niu, S. Siler, and A. S. Gertner: 1998, 'Student Modeling from Conventional Test Data: A Bayesian Approach Without Priors'. In: *ITS'98: Proceedings of the 4th International Conference on Intelligent Tutoring Systems*. London, UK, pp. 434–443.

Vassileva, J.: 1995, 'Dynamic Courseware Generation: at the Cross Point of CAL, ITS and Authoring'. In: *Proceedings International Conference on Computers in Education, Proc. ICCE'95*. pp. 290–297.

Vassileva, J., G. McCalla, , and J. Greer: 2003, 'Multi-Agent Multi-User Modeling in I-Help'. *User Modeling and User-Adapted Interaction* **13**, 179–210.

Villano, M.: 1992, 'Probabilistic Student Models: Bayesian Belief Networks and Knowledge Space Theory'. In: C. e. a. Frasson (ed.): *Proceedings of the second International Conference on Intelligent Tutoring Systems*. pp. 492–498.

Vomlel, J.: 2004, 'Bayesian networks in educational testing'. *International Journal of Uncertainty, Fuzziness and Knowledge Based Systems* **12**, 83–100.

Vygotsky, L. S.: 1978, *Mind in Society: Development of Higher Psychologicial Processes*. Cambridge, MA: Harvard University Press.

Walker, E., S. Walker, N. Rummel, , and K. Koedinger: 2010, 'Using Problem-Solving Context to Assess Help Quality in Computer-Mediated Peer Tutoring'. In: *10th International Conference on Intelligent Tutoring Systems (ITS2010)*.

Walonoski, J. A. and N. T. Heffernan: 2006, 'Detection and Analysis of Off-Task Gaming Behavior in Intelligent Tutoring Systems'. In: M. Ikeda, K. D. Ashley, and T.-W. Chan (eds.): *Intelligent Tutoring Systems, 8th International Conference, ITS 2006, Jhongli, Taiwan, June 26-30, 2006, Proceedings*, Vol. 4053 of *Lecture Notes in Computer Science*. pp. 382–391.

Woolf, B. P.: 2009, *Building Intelligent Interactive Tutors Student-centered strategies for revolutionizing e-learning.* Elsevier.

Xu, Y. and J. Mostow: 2011, 'Using Logistic Regression to Trace Multiple Subskills in a Dynamic Bayes Net'. In: C. Conati, S. Ventura, M. Pechenizkiy, and T. Calders (eds.): *Proceedings of EDM 2011, The 4th International Conference on Educational Data Mining.* p. (to appear).

Yudelson, M., P. I. Pavlik, and K. R. Koedinger: 2011, 'User Modeling – A Notoriously Black Art'. In: *User Modeling, Adaptation, and Personalization, 19th International Conference, UMAP 2011, Gerona, Spain, July 11–15, 2011. Proceedings.* p. (to appear).

Zapata-Rivera, J.-D. and J. Greer: 2004, 'Inspectable Bayesian student modeling servers in multi-agent tutoring systems'. *International Journal of Human-Computer Studies* **61**(4), 535–563.

Authors' Vitae

Michel C. Desmarais

is Assistant Professor at the Computer and Software Engineering Department of Polytechnique Montreal since 2002. He received his Ph.D. degree in psychology in 1990 from the University of Montreal. He was team leader of the HCI and Learning Environments groups at the Computer Research Institute of Montreal between 1990 and 1998, where he was involved in a number of research projects in close collaboration with private corporations. From 1998 to 2002, he directed R&D software development projects in a private company. His research interests are in user modeling, e-learning, human-computer interactions, and software engineering.

Ryan Baker

is Assistant Professor of Psychology and the Learning Sciences in the Department of Social Science and Policy Studies at Worcester Polytechnic Institute, with a collaborative appointment in Computer Science. He is director of WPI's Educational Psychology Laboratory, and is Associate Editor of the Journal of Educational Data Mining. He received his Ph.D. in Human-Computer Interaction at Carnegie Mellon University in 2005. Prior to his current position at WPI, he served as Technical Director of the Pittsburgh Science of Learning Center DataShop, the world's leading public repository for data on the interaction between students and educational software. His research interests include educational data mining, student modeling, intelligent tutoring systems, motivation and affect, and human-computer interaction.